# 用于视频内容认证的抗二次压缩水印算法

付剑晶[1),2)] 陈德人[1)]

[1)](浙江大学计算机科学与技术学院 杭州 310058)

[2)](浙江传媒学院新媒体学院 杭州 310018)

**摘 要** 随着视频编辑软件功能的增强,视频内容的篡改和伪造变得越来越容易,在很多领域可能导致严重后果.在视频依次经历解码、常规信号处理及修改图像组 *GOP*(Group of Pictures)与量化参数 *QP*(Quantization Parameters)的二次压缩条件下,如何有效判定视频内容的真实性是个亟待解决的问题.因此,该文提出了一种基于调制奇异值贡献率的视频内容认证水印算法:(1)通过在视频亮度分量子块的 2 级 DWT(Discrete Wavelet Transform)域抽取特征与嵌入水印,能从整体上提升重构的特征与恢复的水印的一致性;(2)选择在 2 级 DWT 的低频域,基于量化调制第一奇异值贡献率的新颖算法来嵌入水印,实现了水印的透明性和抗二次视频压缩的鲁棒性;(3)基于能量关系所表达的视频内容特征,能在二次压缩条件下保持高度的稳定性;(4)采用水印交叉嵌入法与邻域判定法,提升了算法的安全性及判定的准确性.理论分析表明,提出的算法能抵制伪造、拼贴等攻击,在无恶意篡改的二次压缩条件下,水印的鲁棒性为 0.96.实验结果表明,水印的嵌入对视觉质量的影响平均下降 1.3%,对码率的影响平均增加 8.41%,提出的算法在 H.264/AVC 二次压缩及期间的常规信号处理条件下能有效区分视频保持内容的处理与恶意篡改,并定位篡改位置.

**关键词** 二次压缩;内容认证;视频水印;奇异值分解;贡献率

**中图法分类号** TP309 **DOI 号** 10.11897/SP.J.1016.2018.00558

# Watermarking Algorithm of Tolerating the Second Compression for Video Content Authentication

FU Jian-Jing[1),2)] CHEN De-Ren[1)]

[1)](*College of Computer Science and Technology, Zhejiang University, Hangzhou* 310058)

[2)](*College of New Media, Zhejiang University of Media and Communications, Hangzhou* 310018)

**Abstract** With the function enhancement of video editing software, video contents tampering and forging become more and more easier, which makes the user difficult to distinguish the authenticity of video content. So it may lead to serious consequences when it comes to forensic evidence, medical expertise, military intelligence, national security, and so on. Up to now, most of the existing researches focus on the identification of video content under the first compression or recompression. But after videos are published in compressed form, it often undergo decoding, conventional signal processing, and encoded again by modifying *GOP*(Group of Pictures) and *QP* (Quantization Parameters). Under the secondary compression condition above, how to effectively determine the authenticity of video content is an urgent problem. Therefore, we propose a watermarking algorithm, based on quantitatively modulating the contribution rate of the singular value, for video content authentication. (1) By extracting features and embedding watermark in the 2nd level DWT(Wavelet Transform Discrete) domain for each sub-block from the video luminance component, the consistency between the reconstructed features and extracted watermark can be promoted in the whole; (2) The low frequency domain in the 2nd level DWT is

收稿日期:2016-07-31;在线出版日期:2017-07-28.本课题得到国家自然科学基金(61502415)、浙江省自然科学基金项目(LY18F020003,LY15F020016)、浙江省科技厅公益技术项目(2015C31110)资助.**付剑晶**,男,1976 年生,博士,副教授,主要研究方向为数字水印和数字内容安全等.E-mail:fjjmsn@163.com.**陈德人**,男,1951 年生,教授,博士生导师,主要研究领域为信息安全、电子商务、数据挖掘等.

selected to embed the watermark by using a novel algorithm based on quantitatively modulating the contribution rate of the first singular, which achieves watermark transparency and the robustness tolerating the second video compression. The core content of authentication algorithm based on digital watermarking is the stability of hidden information which is embodied by watermark algorithm. Because the stability of the proposed watermark comes from all the singular values, rather than part of the singular value, algorithm of this paper have obvious advantages in robustness; (3) The video content features based on energy relations can maintain a higher stability after the secondary compression; (4) The security and accuracy of the algorithm are improved respectively by crossing watermark and neighborhood decision. In addition, in this paper, the security of the proposed algorithm is analyzed theoretically, including cryptanalysis, collage attack, quantization attack and targeted forgery attack. And the theoretical analysis shows that the proposed algorithm can resist forgery attack, collage attack and some other attacks, and the robustness of the watermark is 0.96 under the condition of the secondary compression without malicious tampering. Finally, the experiment is carried out aiming at the transparency of watermark, the impact on bit rate and the robustness of attack. And the experimental results show that the influence of watermark embedding on visual quality is decreased by 1.3%, and that on bit rate is increased by 8.41%; that the proposed watermarking scheme can effectively distinguish malicious manipulations from the content-preserving operator after secondary compression for H.264/AVC video, and locate the position tampered. Also, through the performance comparison of related schemes, it further demonstrates the advantage of this algorithm for video content authentication under the condition of secondary compression.

**Keywords**　second compression; content authentication; video watermarking; singular value decomposition; contribution rate

## 1　引　言

随着多媒体处理技术的飞速发展与各种视频编辑软件的出现，视频数据能被轻易篡改和伪造，使其难辨真伪，在涉及到法庭举证、医疗鉴定、军事情报、国家安全等领域时，可能造成严重后果，如何有效地鉴别数字视频内容的真实性得到了广泛关注.

目前，用于数字视频认证的技术主要有传统的数字签名、感知哈希[1]和数字水印技术. 由于传统的数字签名技术对数据中任何比特的改变很敏感，且因签名信息必须单独传送而存在丢弃或篡改的危险，所以在应用中有一定的局限性. 视频感知哈希[2-3]是视频数据集到感知摘要集的一种单向映射，它对视频常规处理不敏感，对恶意攻击很敏感，常应用于多媒体数据检索；但当其用于内容认证时，存在需要单独传送摘要信息的额外开销问题，并且也不能定位被篡改内容的具体位置. 基于数字水印的视频认证技术分为完全级认证与内容级认证，分别对应脆弱与半脆弱数字视频水印技术[4-5]. 采用脆弱水印技术的视频认证不允许对视频数据执行任何更改，否则认证失败；而采用半脆弱视频水印的认证技术，对于常规保持视频图像内容的操作，如有损压缩、格式转换、去噪等，则被认为是可接受的无恶意修改，并对内容的恶意篡改敏感.

然而在实际应用中，人们更关注的是视频内容是否被篡改，定位篡改位置，以及可能的恢复，而非信息在传输、存储及常规加工中少数比特位的误差. 此外，由 ISO 和 ITU-T 联合开发的 H.264/AVC 视频编码标准，以其高效的压缩比率、优质的视频质量及良好的容错能力被得到了广泛应用，但在这个标准中没有涉及到与数据真实性相关的内容，因而适用于 H.264/AVC 的半脆弱水印算法的研究已为该领域的一个重点.

为了兼顾算法的鲁棒性和运行效率，近年来提出的基于 H.264/AVC 的水印算法主要集中在编码中(即编码域). 大多数文献属于面向版权保护的鲁棒性水印算法[6]和面向完整性认证的脆弱水印算

法,而针对内容认证的半脆弱水印算法的研究相对较少.王小静等人[7]利用块组间的能量关系抽取特征码,并将特征码与帧号作为水印嵌入到 DCT 系数中;提出的算法能对时域与空域改进行定位,并对噪声、转码保持内容的处理有一定的鲁棒性.为了增强特征码的鲁棒性,林志高等人[8]提出了一种用于内容认证的半脆弱水印算法.该算法根据块编码模式 CBP(Coded Block Pattern),在自适应选取的宏块中基于能量关系建立了体现内容的特征,并将其嵌入到子宏块拖尾系数中,实现了宏块级别的检测和定位,然而算法存在特征码与水印的干扰问题,且漏检率偏高.文献[9]给出了一种新颖的半脆弱水印方案,该方案利用不同块组间 DCT 系数预测值与残差值之和的能量保持关系抽取特征,再调制 DCT 系数符号统计特征实现水印嵌入,实现了宏块级空域篡改定位;此外将帧索引进行纠错编码作为水印嵌入到 DCT 系数中,以实现时域篡改定位.实验表明该方案能区分恶意攻击与普通信号处理;但由于水印嵌入在 DCT 对角系数的固定位置上,因而会带来安全问题.为此,Farfoura 等人[10]在文献[9]基础上,通过引入安全机制解决水印针对性攻击问题,并通过低复杂度的空域分析,同时提升了水印视频的鲁棒性与视觉质量.

这些研究成果[7-10]为进一步研究视频的内容认证奠定了较好基础,但离实用尚存较大差距:(1)被保护的视频在依次经历常规信号处理与重压缩后,检测失败;(2)视频时域篡改上的检测都是在假定未受任何攻击的前提下进行的,而这种假定在现实中难以成立.

此外,更值得重视的是,视频编码压缩发布后往往要经历如下的应用过程:(1)正常使用,未被加工与再编码;(2)二次应用,视频被解码、帧信号图像增强、再编码;(3)合理使用,为降低传输与存储负荷,视频被解码、下调分辨率等、再编码;(4)盗用,视频被解码、信号处理、再编码;(5)恶意篡改,视频被解码、时域篡改、空域局部内容篡改、信号处理、再编码.而在上述几个过程的再编码中,很可能会对图像组(*GOP*)与量化参数(*QP*)进行有意或无意地修改.其中,过程(1)~(4)中,视频的语义未发生改变,是保持内容的处理,而过程(5)则会改变视频的语义.

在本文,我们称视频经历两次变参编码压缩(同时修改 *GOP* 与 *QP*)及期间的常规信号处理的操作为"二次压缩".由于第二次压缩极可能修改 *GOP*,使得水印嵌入时的 I、B、P 帧与被检测视频的帧类型不能建立起对应关系,从而导致水印嵌入与检测不同步,所以基于编码域和码流域的视频水印方案不能适用于这种二次压缩的应用情形.要针对二次压缩来设计视频内容级认证算法只能考虑放在视频的原始域,这也对算法的鲁棒性提出了很高要求.考虑到视频原始域的每一帧是静态图像,但因图像与视频在编码压缩、性能需求等方面的不同,有关数字图像的内容级认证水印算法不能直接引入到视频领域.到目前,能抵制二次压缩攻击的视频水印文献未见报道,研究面向二次压缩的视频内容级认证算法,已成为该领域一个亟待解决的问题.

视频内容级认证技术包括针对帧图像局部篡改检测的空域认证,和针对帧插入、删除、交换检测的时域认证.就时域认证而言,已有的研究都是将帧索引信息隐藏在帧图像数据块中,检测时根据实际帧索引与提取的帧索引的一致性来判断[6-7,9-10].这种认证思想要求水印算法必须具有近 100% 的稳健性,否则帧索引水印提取的 1 bit 错误就会引起帧序号转换的重大差别,从而导致误判.在保持内容处理的条件下,视频水印算法在满足空域内容认证的同时,如何实现时域上的认证有待研究.

综上所述,现有的视频内容认证水印算法不能区分出二次压缩条件下保持内容的处理与恶意篡改,应用受到很大的局限.基于此,本文提出了一种新颖的视频内容认证水印算法,该算法首先对原始视频每一帧的 $Y$ 分量进行分块 2 级 DWT 变换,并在变换域基于密钥建立数据块特征码,然后将特征水印基于密钥交叉嵌入到其它块的低频子带.在水印嵌入时,本文提出了一种新颖的方法,即量化调制奇异值贡献率.认证时通过比较重构的特征与对应恢复的水印,并结合邻域特征来判定数据块内容的真实性.文中分析了算法的安全性,并推导了算法的检测性能,即鲁棒性、漏警率与虚警率;实验表明,提出的算法在二次 H.264 变参压缩及期间的常规信号处理条件下,能有效区分保持内容处理的操作与恶意篡改攻击,并实现篡改定位.

本文的主要贡献归纳如下:

(1)提出了奇异值贡献率具有稳定性的观点,并给出了理论分析与实验验证,基于此可设计一些新算法应用于多媒体数字水印及感知哈希等应用领域.

(2)提出了基于量化调制第一奇异值贡献率的水印嵌入算法.该算法通过量化方式修改第一奇异值贡献率来嵌入水印信息,并相应调整其余奇异值

贡献率，使得其余奇异值贡献率在调整前后相互间的比例关系保持不变，从而实现嵌入算法的水印透明性和抗二次压缩的鲁棒性.

(3) 在测试数据和理论推导基础之上建立了检测性能度量模型，对算法的实际应用具有指导作用.

本文第 2 节从理论上分析奇异值贡献率的稳定性，并通过实验验证其在二次 H.264 变参压缩条件下的稳定性能；第 3 节阐述特征水印生成算法，然后由经实验获取其在二次 H.264 变参压缩条件下的稳定性数据；第 4 节详细描述本文认证算法，包括奇异值贡献率的量化调制、水印嵌入、特征重建与水印提取、篡改检测与定位；第 5 节分析认证算法的安全性，并在之前的实验数据基础上，推导算法的检测性能度量模型；第 6 节为实验结果与分析；第 7 节对全文做总结.

## 2　奇异值贡献率的稳定性

奇异值贡献率的稳定性是本文认证算法的理论基础. 本节先简要地引入奇异值分解(Singular Value Decomposition，SVD)的相关内容，定义奇异值贡献率，然后从理论上分析奇异值贡献率的稳定性，并进行实验验证.

设矩阵 $\boldsymbol{A}=\{a_{ij}\}_{m\times n}$，且 $a_{ij}\geq 0$，对 $\boldsymbol{A}$ 进行奇异值分解有 $\boldsymbol{A}=\boldsymbol{USV}^{\mathrm{T}}=\sum_{i=1}^{r}\lambda_i U_i \boldsymbol{V}_i^{\mathrm{T}}$. 其中 $\boldsymbol{U}$、$\boldsymbol{V}$ 分别为 $m\times m$ 和 $n\times n$ 的正交矩阵；$\boldsymbol{S}$ 是非负对角阵，$\boldsymbol{S}$ 中的 $\lambda_1,\lambda_2,\cdots,\lambda_r$ 称为矩阵 $\boldsymbol{A}$ 的奇异值，且满足 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_r>0$，其中 $r$ 为 $\boldsymbol{A}$ 的秩. 我们定义矩阵 $\boldsymbol{A}$ 的奇异值 $\lambda_i$ 的贡献率为

$$R_i^{\boldsymbol{A}}=\lambda_i\Big/\sum_{j=1}^{r}\lambda_j \tag{1}$$

设 $\boldsymbol{B}=\boldsymbol{A}+\boldsymbol{E}\in R^{m\times n}$，即 $\boldsymbol{B}$ 为对 $\boldsymbol{A}$ 施加扰动 $\boldsymbol{E}$ 后的矩阵，则 $\boldsymbol{B}$ 的奇异值贡献率相对于 $\boldsymbol{A}$ 具有稳定性，分析过程如下.

分别对 $\boldsymbol{A}$，$\boldsymbol{B}$ 执行 SVD 变换，可得 $\boldsymbol{A}=\boldsymbol{USV}^{\mathrm{T}}=\sum_{i=1}^{r'}\lambda_i U_i \boldsymbol{V}_i^{\mathrm{T}}$，$\boldsymbol{B}=\hat{\boldsymbol{U}}\hat{\boldsymbol{S}}\hat{\boldsymbol{V}}^{\mathrm{T}}=\sum_{i=1}^{r''}\xi_i\hat{\boldsymbol{U}}_i\hat{\boldsymbol{V}}_i^{\mathrm{T}}$，根据 SVD 理论[11]有

$$\|\boldsymbol{S}-\hat{\boldsymbol{S}}\|_2<\varepsilon,\ \varepsilon=\varepsilon(\boldsymbol{E}),$$

故对任意的 $i$，有

$$|\lambda_i-\xi_i|<\varepsilon,$$

$$\Big|\sum\lambda_i-\sum\xi_i\Big|<r_m\varepsilon,$$

其中 $r_m=\max(r',r'')$，因而

$$\frac{\lambda_i-\varepsilon}{\sum\lambda_i+r_m\varepsilon}\leq\frac{\xi_i}{\sum\xi_i}\leq\frac{\lambda_i+\varepsilon}{\sum\lambda_i-r_m\varepsilon},$$

即

$$\frac{\lambda_i-\varepsilon}{\sum\lambda_i+r_m\varepsilon}\leq R_i^{\boldsymbol{B}}\leq\frac{\lambda_i+\varepsilon}{\sum\lambda_i-r_m\varepsilon}.$$

当 $\varepsilon\to 0$ 时，$R_i^{\boldsymbol{B}}\to\lambda_i\Big/\sum_{i=1}^{r_m}\lambda_j$，即 $R_i^{\boldsymbol{B}}\to R_i^{\boldsymbol{A}}$.

以上分析说明对一矩阵 $\boldsymbol{A}$ 施加一个微小抖动，$\boldsymbol{A}$ 的任一奇异值贡献率具有一定的稳定性. 为了验证上述分析结论，并观察矩阵奇异值贡献率稳定性的实际性能，我们进行如下实验.

从标准视频库从选取 5 个 CIF 格式(352×288)视频(Flower，Hall，Akiyo，Stefan，Mobile)，并从每个选取的视频中随机抽取 20 帧组合成一个测试对象. 针对测试对象每帧的 $Y$ 分量，将其划分为互不重叠的子块(16×16)，然后对每个子块执行二级 DWT 变换及低频数据块的 SVD 变换. 我们用奇异值贡献率变化率 $VR_i^{\boldsymbol{A}}$ 来度量数据块 $\boldsymbol{A}$ 的奇异值 $\lambda_i$ 的贡献率的稳定性：

$$\Delta R_i^{\boldsymbol{A}}=|R_i^{\boldsymbol{A}'}-R_i^{\boldsymbol{A}}| \tag{2}$$

$$VR_i^{\boldsymbol{A}}=\Delta R_i^{\boldsymbol{A}}/R_i^{\boldsymbol{A}} \tag{3}$$

其中 $R_i^{\boldsymbol{A}}$ 与 $R_i^{\boldsymbol{A}'}$ 分别表示数据块 $\boldsymbol{A}$ 被攻击前与后对应的 2 级 DWT 低频子带($LL_2$)第 $i$ 个奇异值 $SV_i$ 的贡献率；$\Delta R_i^{\boldsymbol{A}}$ 为攻击前后第 $i$ 个奇异值贡献率的改变量. 测试在 Matlab2011b 环境下进行，选用 H.264/AVC JM10.2 的编/解码器，设计如下 13 种攻击：1-乘积性噪声(0.002)，2-高斯噪声(0.002)，3-椒盐噪声(0.002)，4-滤波均值(5×5)，5-滤波中值(5×5)，6-滤波维纳(5×5)，7-均衡化(164)，8-绝对变暗(−30)，9-绝对增亮(+30)，10-相对变暗([0，0.7])，11-相对增亮([0.4，1])，12-缩放大(1.8×1.8)，13-缩放小(0.8×0.8). 此外，第一次编码主要参数为 $QP_1=24$，$GOP_1=$IPPPPPPPI…；第二次编码主要参数为 $QP_2=28$，$GOP_2=$IBBBPBBBI….

测试过程如图 1 所示，根据是否包含 H.264 二次编解码，分两种测试方案. 数据块受攻击后各奇异值 $SV_i$ 贡献率改变率 $VR_i$ 与累积概率 $CP$ 关系如图 2 所示，其中图(a)为测试方案 1 的结果，即不含 H.264 编解码过程，图(b)为测试方案 2 的结果，包含两次编码与期间的攻击，图(c)是在测试方案 2 的基础上计算 $SV_1$ 的绝对改变量 $\Delta R_1$ 与 $CP$ 的曲线关系. 根据测试数据，可得到以下主要结果：(1) 奇异值 $SV_i$ 贡献率的稳定性随着 $i$ 的增加而降低；

(2)在无 H.264 编码时,$CP(VR_1 \leq 0.2)=0.9902$;(3)在同时修改 $QP$ 与 $GOP$ 参数值的两次 H.264 编码时,$CP(VR_1 \leq 0.2)=0.8053$;(4)两次编码后 $CP(\Delta R_1 \leq 0.2)=0.9465$,$CP(\Delta R_1 \leq 0.1)=0.801$.

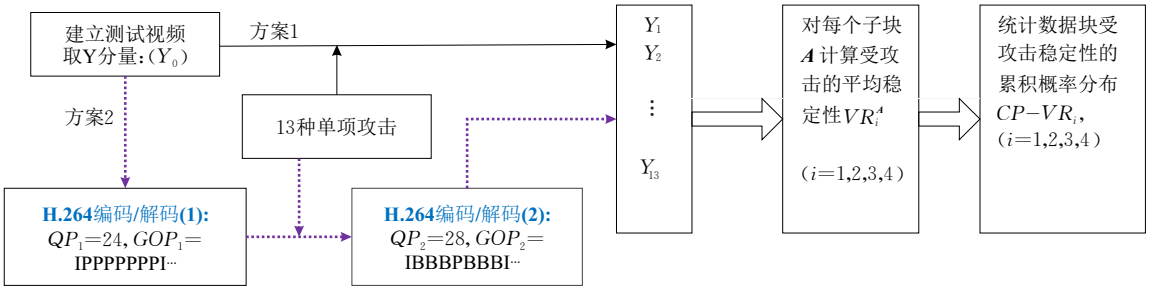
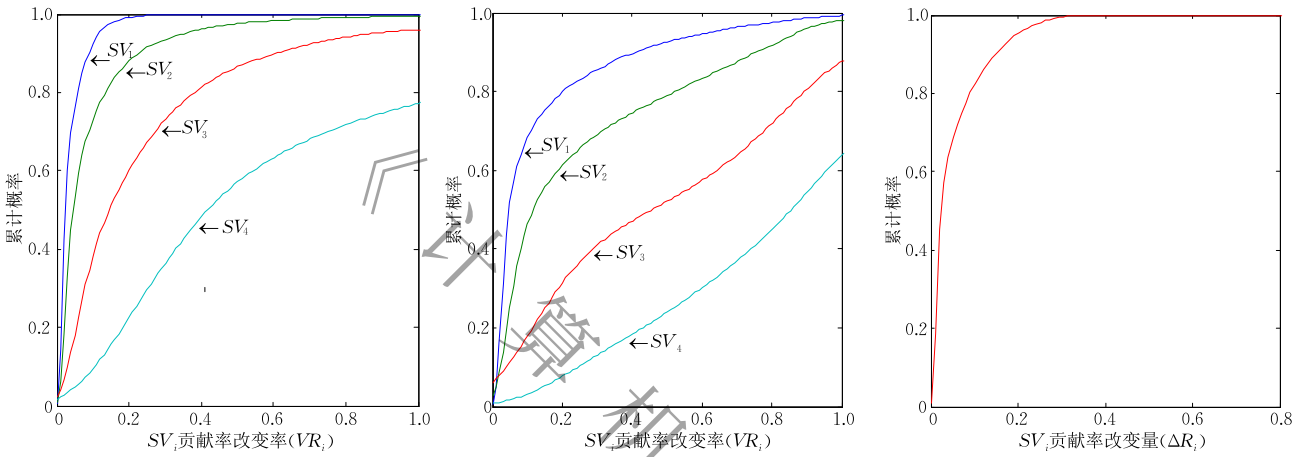



图 1 奇异值贡献率的稳定性测试过程




(a) 无编码时的情形 (b) 包含两次H.264编解码与期间的攻击 (c) 在两次编解码与攻击情况下的$SV_1$的绝对改变量($\Delta R_1$)与累积概率($CP$)的关系

图 2 奇异值贡献率的稳定性

通过以上实验,进一步验证了数据块奇异值贡献率具有一定的稳定性,同时也说明基于第一奇异值贡献率的稳定性来设计可容忍 H.264 两次变参编码压缩的水印算法是可行的.

## 3 特征生成及其稳定性

在视频内容级认证中,代表内容的特征建立非常重要.本节给出在本文方案中视频每帧的特征码生成算法,并测试其在两次变参 H.264 编码压缩及之间的单项攻击下的稳定性.

文献[12]表明,MPEG 视频经一定程度的重复压缩后,两个 DCT 块组间的能量关系具有不变性,并利用了这种关系来刻画视频内容特征.受此启发,本文利用原始视频 $Y$ 分量数据块间的能量关系来抽取视频帧内容特征.为增强特征的稳健性,与降低计算开销,选择数据块 2 级 DWT 后的 $LH_2$ 与 $HL_2$ 域来计算能量值;如此也避免了水印嵌入与特征抽取的相互干扰.特征码生成步骤如下:

(1)划分子块.对原始视频每一帧的 $Y$ 分量,将其划分成互不重叠的子块 $B_i$($i=1,2,\cdots,N$),其中,$B_i$ 的大小与 H.264 中的宏块尺寸(16×16)相同;设视频 $Y$ 分量的分辨率为 $m_1 \times m_2$,则 $N=\mathrm{INT}(m_1/16)\times \mathrm{INT}(m_2/16)$.

(2)计算能量.对第 $i$ 个子块 $B_i$,执行 2 级 DWT 变换,选择变换后的两个高频子带 $LH_2$ 与 $HL_2$,按如下公式计算能量 $E_i$:

$$E_i=\sqrt{\sum_{j=1}^{4}\sum_{k=1}^{4}(C_{j,k}^{h})^2}+\sqrt{\sum_{j=1}^{4}\sum_{k=1}^{4}(C_{j,k}^{v})^2},$$

其中 $C_{j,k}^{h}$ 与 $C_{j,k}^{v}$ 分别表示 $LH_2$ 与 $HL_2$ 的系数.

(3)生成参考块索引.基于密钥 $K_1$ 建立一维随机向量 $\boldsymbol{R}=[r_1,r_2,\cdots,r_N]$,对 $\boldsymbol{R}$ 排序,得到索引向量 $\boldsymbol{C}=[c_1,c_2,\cdots,c_N]$,其中 $c_i$,为 1 到 $N$ 的整数,$i=1,2,\cdots,N$.

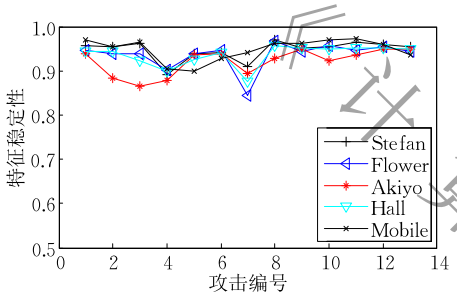
(4)计算数据块的特征码.按如下公式计算数据块 $B_i$ 的 1 比特特征码 $f_i$

$$f_i=\begin{cases}1, & E_i \geq E_{c_i}\\ 0, & E_i < E_{c_i}\end{cases}.$$

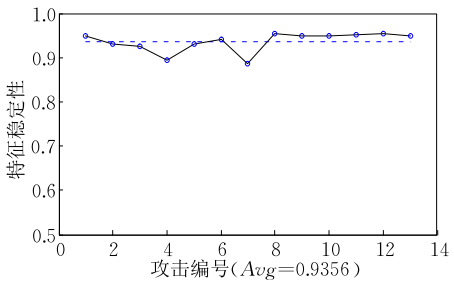
(5) 建立单帧特征码序列. 对原始视频每一帧的 $Y$ 分量，根据上述步骤建立特征码序列

$$F=(f_1,f_1,\cdots,f_N).$$

为观察上述算法生成的特征码经攻击后的实际稳定性，我们执行了一系列测试，测试环境与上一节完全相同，测试结果图 3 所示. 图 3(a)给出了测试对象中 5 个视频片段分别受攻击的特征稳定性情况，数据范围在 0.85 到 0.97 之间，其中 Flower 视频片段在两次编码加攻击(7-均衡化(164))时达到最小值. 稳定性不小于 0.94 与 0.90 的数据块分别占 60%与 87%. 图 3(b)反映了测试对象总体特征的稳定性，数据范围在 0.89 到 0.96 之间，稳定性不小于 0.94 与 0.90 的分别占 69.2%与 92.3%；数据块特征受两次编码及单项攻击的平均稳定性为 0.9356.




(a) 测试对象中5个视频片段各自受攻击的特征稳定性




(b) 测试对象总体特征平均稳定性

图 3　特征经受两次变参编码及期间的 13 种单项攻击后的稳定性

## 4　认证算法

为了区分对视频内容的恶意篡改与无恶意信号处理，本文要建立一种主动认证算法，即把代表视频内容的特征码作为辅助信息(水印)嵌入到视频载体中；在检测时，按照给定的策略，并基于重建的特征码与被提取出来水印的一致性来判定内容的真实性. 由于在实际应用中，为了达到对视频的盗用、内容的恶意篡改、无恶意信号处理、降低带宽等目的，需要对视频进行二次压缩，即解码、信号处理、再编码操作，而且在第二次编码压缩时极可能修改 GOP 参数，因而水印认证算法只能是前置式的(基于视频原始域)，而且水印嵌入算法抗视频二次压缩的鲁棒性也成为了整个认证算法的关键.

在前两节的基础上，本节针对原始视频每一帧的 $Y$ 分量，通过量化调制第一奇异值贡献率来实现特征水印的嵌入；在接收端，根据相应的密钥就可以重建特征及提取水印，从而实现视频各帧的内容认证. 本节先描述基于第一奇异值贡献率的量化调制算法，然后给出了水印嵌入与提取的具体过程，最后对视频空域内容篡改的检测与定位进行了详细阐述.

### 4.1　奇异值贡献率的量化与调整

设矩阵 $\boldsymbol{A}=\{a_{ij}\}_{m\times n}$，其奇异值为 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_r>0$，其中 $r$ 为 $\boldsymbol{A}$ 的秩. 当 $r\leq 1$ 时，$\boldsymbol{A}$ 为零矩阵数据块或镜面数据块，需进行预处理：(1) 当 $r=0$ 时，令 $\lambda_1=10^{-5}$，$\lambda_2=10^{-9}$；(2) 当 $r=1$ 时，令 $\lambda_2=\lambda_1\times 10^{-5}$. 根据式(1)，各奇异值贡献率和为 1，且有 $R_1^{\boldsymbol{A}}>1/r$，其中 $r>1$. 为了在 $\boldsymbol{A}$ 中嵌入 1 bit 特征水印 $f_w$，先量化调整第一奇异值贡献率 $R_1^{\boldsymbol{A}}$，然后相应修改其余奇异的值贡献率，由此可实现各奇异值的调整，最后执行 SVD 反变换.

设量化步长为 $q\in(0,1)$，对 $R_1^{\boldsymbol{A}}$ 的量化方法如图 4 所示. 将数轴划分为两类区间，即$[(k-1)q,kq)$与$[(2k-1)q,2kq)$，其中 $k$ 为不小于 1 的整数. 当嵌入的 $f_w$ 为 0、1 时，分别将 $R_1^{\boldsymbol{A}}$ 调整到两类区间的起点，记为 $R_1^{\boldsymbol{A}'}$ 且满足 $R_1^{\boldsymbol{A}'}>1/r$，$R_1^{\boldsymbol{A}'}$ 计算方法如下：

$$R_1^{\boldsymbol{A}'}=\begin{cases} floor(R_1^{\boldsymbol{A}}/q)\times q, & floor(R_1^{\boldsymbol{A}}/q)\times q>1/r \ \&\& \\ & floor(R_1^{\boldsymbol{A}}/q)==f_w \\ floor(R_1^{\boldsymbol{A}}/q)\times q+2q, & floor(R_1^{\boldsymbol{A}}/q)\times q\leq 1/r \ \&\& \\ & floor(R_1^{\boldsymbol{A}}/q)==f_w \\ floor(R_1^{\boldsymbol{A}}/q)\times q+q, & floor(R_1^{\boldsymbol{A}}/q)\times q+q>1/r \ \&\& \\ & floor(R_1^{\boldsymbol{A}}/q)\neq f_w \\ floor(R_1^{\boldsymbol{A}}/q)\times q+3q, & floor(R_1^{\boldsymbol{A}}/q)\times q+q\leq 1/r \ \&\& \\ & floor(R_1^{\boldsymbol{A}}/q)\neq f_w \end{cases},$$

其中 $floor(\cdot)$为向下取整操作. 当 $R_1^{\boldsymbol{A}'}>1$ 时，进一步调整为 $R_1^{\boldsymbol{A}'}=R_1^{\boldsymbol{A}'}-2q$.

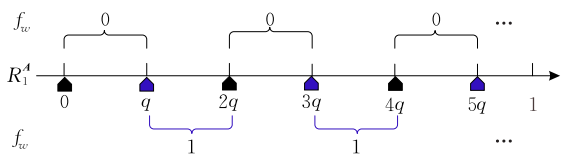



图 4　第一奇异值贡献率量化规则

接下来考虑其余奇异值贡献率的调整，即调整 $R_j^{\boldsymbol{A}}$ 到 $R_j^{\boldsymbol{A}'}$ $(j=2,3,\cdots,r)$. 为减小奇异值贡献率调整后对视频视觉质量的影响，调整方法作如下约束：(1)各奇异值贡献率单调递减，且属于区间[0,1)；(2)非零奇异值贡献率总数保持前后一致；(3)除第一奇异值贡献率外，其余各奇异值贡献率间的比例关系保持前后一致. 如此调整可以保证水印嵌入与检测同步，而且使数据块 $\boldsymbol{A}$ 的主元信息(第一主元除外)在各自分布方向上的比例前后也保持一致[13]. 为此，设 $factor=(1-R_1^{\boldsymbol{A}'})/(1-R_1^{\boldsymbol{A}})$，则其余奇异值贡献率的计算规则定义为

$$R_j^{\boldsymbol{A}'}=R_j^{\boldsymbol{A}}\times factor,\ j=2,3,\cdots,r,$$

其中 $R_j^{\boldsymbol{A}}$ 为修改前第 $j$ 个奇异值对应的贡献率，$factor$ 为贡献率缩放因子. 由此可得，当 $i\neq k$，且 $i,k=2,3,\cdots,r$ 时：(1) $R_i^{\boldsymbol{A}}/R_k^{\boldsymbol{A}}=R_i^{\boldsymbol{A}'}/R_k^{\boldsymbol{A}'}$；(2)若 $R_i^{\boldsymbol{A}}=0$，则 $R_i^{\boldsymbol{A}'}=0$；(3)若 $R_i^{\boldsymbol{A}}>R_k^{\boldsymbol{A}}$，有 $R_i^{\boldsymbol{A}'}>R_k^{\boldsymbol{A}'}$. 为确保 $R_1^{\boldsymbol{A}'}>R_2^{\boldsymbol{A}'}$，即 $R_1^{\boldsymbol{A}'}>R_2^{\boldsymbol{A}'}\times(1-R_1^{\boldsymbol{A}'})/(1-R_1^{\boldsymbol{A}})$，必须满足以下条件：

$$R_1^{\boldsymbol{A}'}>R_2^{\boldsymbol{A}}/(1-R_1^{\boldsymbol{A}}+R_2^{\boldsymbol{A}}).$$

若 $R_1^{\boldsymbol{A}'}\leq R_2^{\boldsymbol{A}}/(1-R_1^{\boldsymbol{A}}+R_2^{\boldsymbol{A}})$，则以 $2q$ 为步长，上调 $R_1^{\boldsymbol{A}'}$ 直到满足条件为止，再重新计算 $R_j^{\boldsymbol{A}'}$ $(j=2,3,\cdots,r)$.

最后计算调整后的各奇异值，即调整 $\lambda_j$ 到 $\lambda_j'$ $(j=1,2,\cdots,r)$：

$$\lambda_j'=R_j^{\boldsymbol{A}'}\times\sum_{i=1}^{r}\lambda_i,$$

其中 $\lambda_j$、$\lambda_j'$ 分别对应数据块修改前后的奇异值；然后通过 SVD 反变换得到嵌入 1bit 信息的数据块 $\boldsymbol{A}'$.

### 4.2 水印的嵌入、提取与特征重建

本节在第 2、第 3 节及 4.1 节的基础上，阐述基于量化调制第一奇异值贡献率的水印方法，包括水印的嵌入与提取，及特征重建. 水印嵌入的主要思想是：对原始 YUV 视频每帧的 $Y$ 分量进行分块 2 级 DWT 变换，在 DWT 域完成特征抽取与水印嵌入，每个数据块嵌入的特征水印来自其它数据块以防止拼贴与伪造攻击. 水印嵌入过程如图 5 所示.

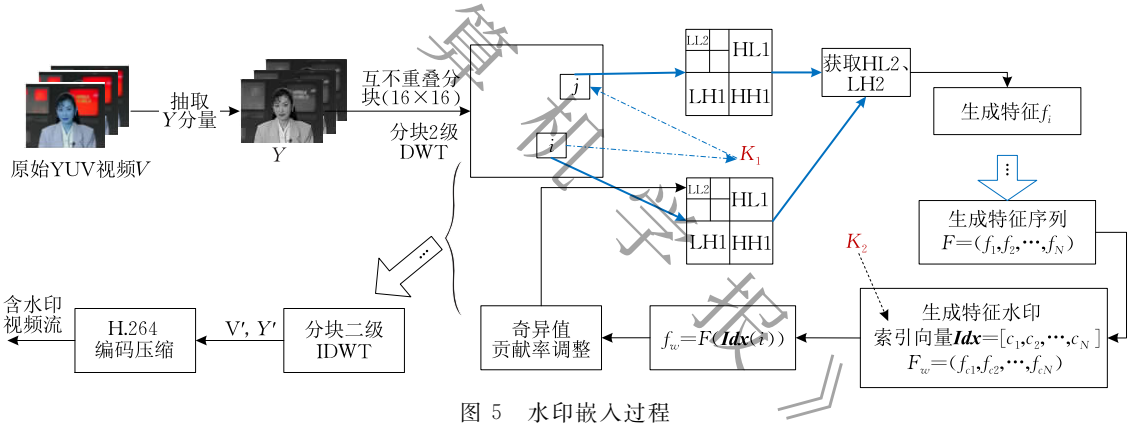



图 5 水印嵌入过程

#### 4.2.1 水印嵌入

(1)按照 H.264 中的宏块尺寸(16×16)，对原始 YUV 视频中每帧的亮度分量 $Y$ 划分为互不重叠的数据子块，记为 $B_i$ $(i=1,2,\cdots,N)$，其中 $Y$ 的分辨率为 $m_1\times m_2$，$N=\mathrm{INT}(m_1/16)\times\mathrm{INT}(m_2/16)$.

(2)为降低亮度的绝对调整对水印稳定性的影响，对 $Y$ 中的每个数据块 $B_i$ 先进行标准化预处理，即令 $avg\_B_i=avg(B_i)$，$B_i=B_i-avg\_B_i$，其中 $avg(\cdot)$ 为计算矩阵均值操作. 然后对 $B_i$ 执行二级 DWT 分解，$B_i$ 的 4 个二级子带分别表示为 $B_{i,LL_2}$，$B_{i,HL_2}$，$B_{i,LH_2}$，$B_{i,HH_2}$；再根据第 3 节的内容，在 $B_{i,HL_2}$，$B_{i,LH_2}$ 子带上建立单帧特征码序列 $F=(f_1,f_2,\cdots,f_N)$.

(3)为了提高水印的安全性，对 $F$ 建立基于密钥 $K_2$ 的索引向量 $\boldsymbol{Idx}=[c_1,c_2,\cdots,c_N]$，实现特征码的置乱. 置乱后的特征码序列形成特征水印，记为 $F_w=(f_{c_1},f_{c_2},\cdots,f_{c_N})$，因此在数据块 $B_i$ 中待嵌入的水印表示为 $f_w=F_w(i)=F(\boldsymbol{Idx}(i))$，表明是来自其它数据块的特征码.

(4)在 $B_i$ 中嵌入 1bit 特征水印. 令 $\boldsymbol{A}=B_{i,LL_2}$，$f_w=F_w(i)$，再根据 4.1 节的方法完成单个数据块的水印嵌入.

(5)反复执行第(4)步，直到每个数据块都嵌入水印.

(6)对 $Y$ 中的每个数据块 $B_i$ 先执行二级 IDWT，然后令 $B_i=B_i+avg\_B_i$，如此可得到含水印信息的单帧亮度分量 $Y'$.

(7)重复执行上述(1)～(6)步，得到含水印的 YUV 视频 $V'$，再经 H.264 压缩编码形成可发布的视频码流文件.

4.2.2　水印提取

水印提取是水印嵌入的逆过程，本算法实现对水印的盲性提取，即不需要原始视频的参与. 算法框架可参照水印嵌入过程，在此只描述对第 $i$ 个数据块 $B_i$ 的 1 bit 水印提取过程：

(1) 令 $\boldsymbol{A}'=B_{i,LL_2}$，对 $\boldsymbol{A}'$ 执行 SVD 分解；

(2) 计算第一奇异值值贡献率 $R_1^{\boldsymbol{A}'}$；

(3) 以量化步长 $q$ 对 $R_1^{\boldsymbol{A}'}$ 进行反量化

$$f_{w'}=mod(round(R_1^{\boldsymbol{A}'}/q),2);$$

其中 $round(\cdot)$、$mod(\cdot)$ 分别为四舍五入与取余操作，其规则如图 6 所示.

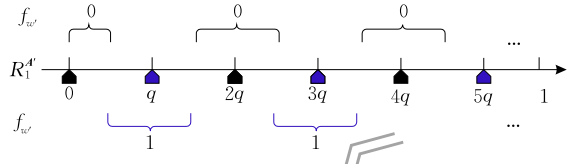



图 6　第一奇异值贡献率反量化规则

按上述提取算法，将原始 YUV 视频中每帧的亮度分量 $Y$ 提取的水印记为 $W'_k=(f_{w'_1},f_{w'_2},\cdots,f_{w'_N})$，再根据密钥 $K_2$ 对 $W'_k$ 执行反置乱，得到各子块对应的水印序列，表示为 $W'=(f_{1'},f_{2'},\cdots,f_{N'})$.

4.2.3　特征重建

根据第 3 节计算步骤(1)～(5)，计算重建的特征码序列，记为 $F'=(f_{G1'},f_{G2'},\cdots,f_{GN'})$，此过程可以集成在水印提取过程中以提高检测效率.

### 4.3　篡改检测与定位

被保护的视频在经水印加载与编码压缩后发布，之后的应用过程可归纳为 5 类(详见第 1 节). 本文的篡改检测与定位的目标是要将过程(1)～(4)的保持内容的处理与过程(5)的恶意篡改进行区分识别，并定位篡改位置.

正如第 1 节引言中所述，在视频受到攻击后，对基于水印的认证方案解决视频时域上的篡改检测与定位，目前还没有效的办法. 本节主要阐述视频经再编码及期间的常规信号处理后，在空域上对内容的篡改进行检测与定位.

设被解码的待测视频每一帧 $Y$ 分量的分辨率为 $m_1\times m_2$，根据 4.2 节算法能得到每个 $16\times16$ 子块 $B_i$ 对应的水印序列与特征码序列，即 $W'=(f_{1'},f_{2'},\cdots,f_{N'})$、$F'=(f_{G1'},f_{G2'},\cdots,f_{GN'})$. 为了表达上的更直观，将提取的水印序列与特征码序列分别转换成二维形式，方法如下：

$$\begin{cases}W''(i,j)=W'(k), & i=(k-1)/(m_2/16)+1\\ F''(i,j)=F'(k), & j=(k-1)\%(m_2/16)+1\end{cases},$$

其中，$1\le i\le m_1/16$，$1\le j\le m_2/16$，$1\le k\le N$. 篡改检测与定位包含以下步骤：

(1) 特征比较. 对每个数据块，比较其提取的水印与重建的特征并生成初始篡改矩阵 $\boldsymbol{T}^0=\{t^0_{i,j}\mid 1\le i\le m_1/16,1\le j\le m_2/16\}$，

$$t^0_{i,j}=\begin{cases}0, & W''(i,j)=F''(i,j)\\ 1, & W''(i,j)\ne F''(i,j)\end{cases}.$$

(2) 邻域不一致块统计. 根据初始篡改矩阵 $\boldsymbol{T}^0$，生成邻域特征矩阵 $\boldsymbol{\Delta}=\{\delta_{i,j}\mid 1\le i\le m_1/16,1\le j\le m_2/16\}$，其中 $\delta_{i,j}=\sum t^0_{k,l}$，$k=i,i\pm1$，$l=j,j\pm1$.

(3) 篡改检测. 根据矩阵 $\boldsymbol{\Delta}$ 修改初始篡改矩阵 $\boldsymbol{T}^0$ 生成篡改矩阵 $\boldsymbol{T}=\{t_{i,j}\mid 1\le i\le m_1/16,1\le j\le m_2/16\}$，

$$t_{i,j}=\begin{cases}1, & \delta_{i,j}\ge5\\ 0, & \text{其他}\end{cases}\tag{4}$$

(4) 篡改评估. 由于待检测视频往往经历了信号处理与再编码的无恶意操作，帧数据发生变化是全局性的，因此篡改矩阵 $\boldsymbol{T}$ 中的 1 分布是以稀疏均匀分布为主，并可能存在少数小半径的集中分布；如果帧内容局部遭到恶意篡改，矩阵 $\boldsymbol{T}$ 中的 1 还会出现较大半径的集中分布. 因此直接使用篡改矩阵 $\boldsymbol{T}$ 不能从客观上有效区分视频的合理使用与恶意篡改，为此我们定义比特错误率 $BER$(Bit Error Rate)作为视频是否遭恶意篡改的客观判定准则：

$$BER=\frac{256\times\sum_{i=1}^{m_1/16}\sum_{j=1}^{m_2/16}t_{i,j}}{m_1\times m_2}\tag{5}$$

(1) 当 $BER\le\tau_1$ 时，视频可能经受常规信号处理与再编码，帧内容真实；

(2) 当 $\tau_1<BER\le\tau_2$ 时，视频经受常规信号处理与再编码，帧内容基本可信；

(3) 当 $BER>\tau_2$ 时，视频遭受内容篡改与再编码，帧内容不可靠.

其中 $\tau_1$、$\tau_2$ 为检测阈值，其理论值可参考 5.2 节中的性能指标 $P_{r|T}$，具体值由使用者根据实际应用而定；当出现(2)、(3)情形时用篡改矩阵 $\boldsymbol{T}$ 中的“1”来标明检测帧对应的篡改位置，以进一步人工识别.

## 5　算法分析

### 5.1　安全性分析

(1) 密码分析攻击

本文认证算法由基于置乱算法的密钥 $K_1$、$K_2$

的控制，其中 $K_1$ 用于特征码的生成，$K_2$ 用于水印嵌入时特征码的交叉选择. 由于置乱算法的密钥空间非常大，在未知密钥的情况下攻击者要破解此系统必须尝试所有的置乱可能，即视频帧数据块编号的排列. 对分辨率为 $m_1 \times m_2$ 的视频帧，本文两个密钥的遍历空间为 $((\mathrm{INT}(m_1/16) \times \mathrm{INT}(m_2/16))!)^2$.

(2) 拼贴攻击

视频拼贴攻击是对由相同密钥和相同算法生成的分辨率一致的两个含水印的视频 $V_a$ 与 $V_b$，将 $V_a$ 某帧一局部区域替换到 $V_b$ 某帧相同位置区域得到内容被篡改的帧. 此类攻击是对视频原始域水印算法的最大威胁之一，现有很多鲁棒性水印算法不能有效抵抗该类攻击. 由于在本文水印算法中，数据块 ***A*** 嵌入的特征来自被密钥 $K_2$ 控制的数据块 ***B***，导致在被拼贴的区域，水印与其特征不能保证同步，因而不能实现对检测端的欺骗.

(3) 量化攻击

根据文献[14]量化攻击的观点，如果帧中每个数据块嵌入的水印与其它数据块的内容无关，而且嵌入算法中未采用加密机制，于是，只要两个数据块嵌入的水印相同，就可以交换二者内容而不会导致认证失败. 由于本文算法的特征来自于数据块内容，并作为水印按交叉方式嵌入到其它数据块中，因此可以抵抗量化攻击.

(4) 针对性的伪造攻击

此类攻击的目标是既要有针对性地篡改数据内容，又要逃避检测. 对于本文算法，攻击者只能按以下途径执行篡改：① 使伪造块的水印与被替换块的一致；② 使伪造块的特征数据与被替换块的一致；③ 同时调整伪造块的水印与特征数据，使之与被替换块都一致.

若使伪造块的水印与被替换块相同，根据 4.1 节算法，在不知量化步长 $q$ 的前提下，则伪造块与被替换块的第一奇异值贡献率必须相同，导致平均有一半的伪造块的第一奇异值贡献率的修改量很大，如此会使这些块出现明显的分块痕迹. 若使伪造块的特征数据与被替换块相同，则必须使伪造块的 $LH_2$ 与 $HL_2$ 域数据与被替换块相同(见第 3 节)，如此又违背了伪造块原来的语义. 此外，由于本文算法中数据块嵌入的水印是来自其它数据块的特征，即水印交叉嵌入，因而篡改方法①产生的伪造块虽然可逃避检测，但会造成半数伪造块出现明显的分块痕迹，而且导致交叉对应的其它数据块半数检测失败；篡改方法②会使伪造块实现不了原本的语义；篡改方法③会同时出现方法①、②带来的问题. 所以针对性的伪造只能采用方法①，对此只要把本文算法的量化步长 $q$ 作为密钥之一，当伪造数据块数超过 $2\tau_2$ 时(见 4.3 节)，则不能逃避认证端的检测.

### 5.2 保持内容处理的检测性能分析

本节对算法的保持内容处理的检测性能进行理论分析，即基于式(4)分析视频帧的单个数据块 $A_{i,j}$ 在处理后的鲁棒性(记为 $P_{r|G}$)、虚警率(真实数据块被判为错误数据块的概率，记为 $P_{fr|G}$)、漏警率(错误数据块被判为真实数据块的概率，记为 $P_{fa|G}$). 需要说明的是，本文算法保持内容的处理是指对含水印的原始视频依次执行编/解码、视频帧常规信号处理、变参编码压缩. 设单个数据块 $A_{i,j}$ 在攻击前后其水印与特征的稳定性分别为 $p_w$、$p_f$，并用 $w$ 与 $w'$，$f$ 与 $f'$ 分别表示 $A_{i,j}$ 攻击前与后的水印及特征，那么抽取的水印与重构的特征不一致的概率为

$$\begin{aligned}P\{w' \neq f'\} &= 1 - P\{w' = f'\} \\ &= 1 - (P\{(w = w') \cap (f = f') \cap (w = f)\} + \\ &\quad P\{(w \neq w') \cap (f \neq f') \cap (w' = f')\}) \\ &= 1 - (p_w \times p_f + (1 - p_w) \times (1 - p_f)) \end{aligned} \tag{6}$$

根据式(4)推导如下：

$$\begin{aligned}P_{r|G} &= P\{t_{i,j} = 0\} = P\{\delta_{i,j} < 5\} \\ &= \sum_{k=0}^{5-1} \binom{9}{k} (P\{w' \neq f'\})^k (1 - P\{w' \neq f'\})^{9-k} \end{aligned} \tag{7}$$

$$\begin{aligned}P_{fr|G} &= P\{t_{i,j} = 1 \mid A_{i,j} \text{为真实块}\} \\ &= P\{\delta_{i,j} \geq 5\} \times P\{A_{i,j} \text{为真实块}\} \\ &= \left(1 - \sum_{k=0}^{5-1} \binom{9}{k} (P\{w' \neq f'\})^k (1 - P\{w' \neq f'\})^{9-k}\right) \times 1 \\ &= 1 - P_{r|G} = BER \end{aligned} \tag{8}$$

$$\begin{aligned}P_{fa|G} &= P\{t_{i,j} = 0 \mid A_{i,j} \text{为错误块}\} \\ &= P\{\delta_{i,j} \leq 5 - 1\} \times P\{A_{i,j} \text{为错误块}\} \\ &= \left(\sum_{k=0}^{5-1} \binom{9}{k} (P\{w' \neq f'\})^k (1 - P\{w' \neq f'\})^{9-k}\right) \times 0 \\ &= 0 \end{aligned} \tag{9}$$

以上推导是在假定保持内容处理的前提下进行的，也就是不存在对内容的恶意篡改，所以 $A_{i,j}$ 为错误块的概率为 0，因而 $P_{fa|G} = 0$. 由于本文算法面对的攻击主要包括两次变参编码压缩及期间的多种信号处理，其中编码压缩涉及众多参数，过程非常复杂，而且不同的信号处理强度受各自参数的影响，所以难以建立 $p_w$ 与 $p_f$ 的解析表达. 为此，我们以本文第 2 节、第 3 节的测试数据为基础，建立二者的参考值. 根据第 2 节式(2)及图 6 所示的反量化规则可得

$$p_w=\sum_k P\{2kq\le\Delta R_1^A\le 2kq+q/2\}$$
$$k=0,1,2,3,\cdots,\text{且 } 2kq+q/2\le 1 \tag{10}$$

当 $q=0.2$ 时，

$$p_w=P\{0\le\Delta R_1^A\le 0.1\}+P\{0.4\le\Delta R_1^A\le 0.5\}.$$

根据形成图 2(c)的测试数据有

$$\begin{aligned}p_w&=CP(0.1)+CP(0.5)-CP(0.4)\\&=0.801+1-1\\&=0.801\end{aligned} \tag{11}$$

再根据第 3 节的测试数据可得

$$p_f=0.9356 \tag{12}$$

把式(11)、(12)代入式(6)，然后计算式(7)、(8)有

$$P_{r|G}=0.96,\ P_{fr|G}=0.04.$$

### 5.3　恶意篡改的检测性能分析

要对本文算法保护的视频内容进行恶意篡改，攻击者必须先解码含水印的 H.264 视频流，然后可能执行常规信号处理以掩饰篡改痕迹，最后执行可能同时改变 *GOP* 与 *QP* 参数的再编码. 本节对上述恶意篡改过程的检测进行理论分析，刻画单个数据块 $A_{i,j}$ 在篡改后的鲁棒性(记为 $P_{r|T}$)、虚警率(记为 $P_{fr|T}$)、漏警率(记为 $P_{fa|T}$). 设 $R_0$、$R_1$ 分别表示视频每帧真实数据块与篡改数据块的集合，$p_T$ 为篡改比例，$p_T\in[0,1]$. 则数据块 $A_{i,j}$ 位于篡改区域内、外的概率为

$$\begin{aligned}&P\{A_{i,j}\in R_1\}=p_T,\\&P\{A_{i,j}\in R_0\}=1-p_T\end{aligned} \tag{13}$$

另设 $A_{i,j}$ 在保持内容处理的前后其水印与特征的稳定性分别为 $p_w$、$p_f$，并用 $w$ 与 $w_t'$，$f$ 与 $f_t'$ 分别表示 $A_{i,j}$ 在篡改攻击前与后的水印及特征. 显然，(1) 当 $A_{i,j}$ 位于篡改区时，无论其特征所在数据块位于真实区还是篡改区，此时抽取的水印与重构特征一致性的概率均为 0.5；(2) 当 $A_{i,j}$ 及其特征所在数据块同时位于真实区时，抽取的水印与重构特征一致性的概率可按式(6)计算；(3) 当 $A_{i,j}$ 位于真实区时，而其特征所在数据块位于篡改区时，此时一致性概率为 0.5. 由此可得

$$\begin{aligned}P\{w_t'\ne f_t'|A_{i,j}\in R_0\}&=1-P\{w_t'=f_t'|A_{i,j}\in R_0\}\\&=1-(P\{w_t'=f_t'|w_t'\in R_0,f_t'\in R_0\}+\\&\quad P\{w_t'=f_t'|w_t'\in R_0,f_t'\in R_1\})\\&=1-(((p_w\times p_f+(1-p_w)\times(1-p_f))\times\\&\quad(1-p_T)+0.5\times p_T),\\P\{w_t'\ne f_t'|A_{i,j}\in R_1\}&=1-P\{w_t'=f_t'|A_{i,j}\in R_1\}\\&=1-0.5\\&=0.5\end{aligned} \tag{14}$$

因此，篡改后水印与特征全局不一致性概率可定义为

$$\begin{aligned}P\{w_t'\ne f_t'\}=&P\{w_t'\ne f_t'|A_{i,j}\in R_0\}\times(1-p_T)+\\&P\{w_t'\ne f_t'|A_{i,j}\in R_1\}\times p_T\end{aligned} \tag{15}$$

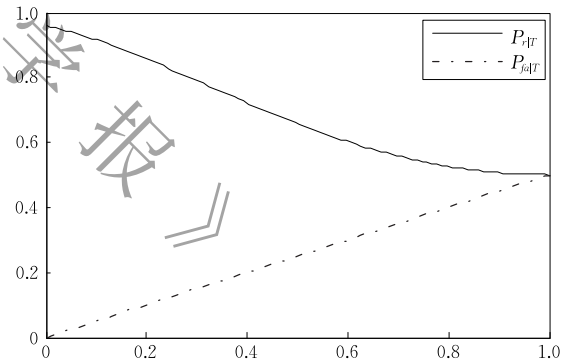
根据式(7)～(9)，同理可得到以下推导：

$$\begin{aligned}P_{r|T}&=P\{t_{i,j}=0\}=P\{\delta_{i,j}<5\}\\&=\sum_{k=0}^{5-1}\binom{9}{k}(P\{w_t'\ne f_t'\})^k(1-P\{w_t'\ne f_t'\})^{9-k}\end{aligned} \tag{16}$$

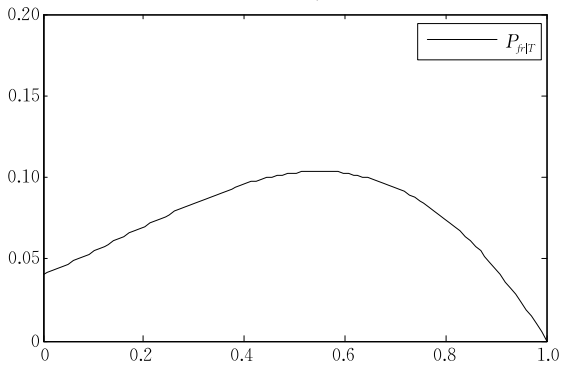
$$\begin{aligned}P_{fr|T}&=P\{t_{i,j}=1|A_{i,j}\text{为真实块}\}\\&=P\{\delta_{i,j}\ge 5|A_{i,j}\in R_0\}\\&=\Big(1-\sum_{k=0}^{5-1}\binom{9}{k}(P\{w_t'\ne f_t'|A_{i,j}\in R_{0t}\})^k\times\\&\quad(1-P\{w_t'\ne f_t'|A_{i,j}\in R_0\})^{9-k}\Big)\times\\&\quad(1-p_T)\end{aligned} \tag{17}$$

$$\begin{aligned}P_{fa|T}&=P\{t_{i,j}=0|A_{i,j}\text{为错误块}\}\\&=P\{\delta_{i,j}\le 5-1|A_{i,j}\in R_1\}\\&=\Big(\sum_{k=0}^{5-1}\binom{9}{k}(P\{w_t'\ne f_t'|A_{i,j}\in R_1\})^k\times\\&\quad(1-P\{w_t'\ne f_t'|A_{i,j}\in R_1\})^{9-k}\Big)\times p_T\\&=\Big(\sum_{k=0}^{5-1}\binom{9}{k}(P\{w_t'\ne f_t'|A_{i,j}\in R_1\})^9\Big)\times p_T\end{aligned} \tag{18}$$

根据式(11)、(12)与式(14)～(18)，经编程计算可描绘篡改比例 $p_T$ 分别与 $P_{r|T}$、$P_{fr|T}$、$P_{fa|T}$ 曲线关系，如图 7 所示，当 $p_T=0$ 时，即保持内容的处理攻




(a) $p_T$与$P_{r|T}$、$P_{fa|T}$的关系




(b) $p_T$与$P_{fr|T}$的关系

图 7　篡改比例 $p_T$ 分别与鲁棒性($P_{r|T}$)、虚警率($P_{fr|T}$)、漏警率($P_{fa|T}$)曲线关系

击，$P_{r|T}=0.96$，$P_{fa|T}=0$，$P_{fr|T}=0.04$；当 $p_T=0.3$ 时，$P_{r|T}=0.7948$，$P_{fa|T}=0.1465$，$P_{fr|T}=0.0832$.

## 6 实验与分析

为测试本文认证算法的主要性能，我们在 Matlab2011b 环境下集成 H.264/AVC JM10.2 的编/解码器进行实验. 选取 5 个 QCIF 格式(176×144，YUV 4:2:0)标准视频序列(News、Mobile、Silent、Coastguard、Hall)作为测试对象；第一次编码采用 Baseline 配置文件，主要参数为 $QP_1=26$，$GOP_1=$IPPPPPPI…，记为 H.264-E1；第二次编码采用 Main 配置文件，主要参数为 $QP_2=28$，$GOP_2=$IBBBPBBBI…，记为 H.264-E2；其它参数采用默认设置. 算法中的量化参数 $q$ 设置为 0.2. 此外，在 H.264-E1 与 H.264-E2 之间采用的保持内容处理的攻击请参考本文第 2 节. 由于视频主要是以编码压缩方式来存储与传播的，本文实验选择在 H.264-E1 后对有/无水印的视频进行相关性能比较.

### 6.1 水印对视觉质量的影响

为了评价视频加载水印后对视觉质量的影响，图 8、图 9 分别展示了原始视频与含水印的视频分别在 H.264-E1 后第 15 帧的图像. 对照图 8、图 9，从主观视觉感知上难以分辨重要差别；通过对测试视频前 100 帧的检测发现，除了在极少数近镜面数据块存在轻微的人工痕迹外，其它区域观察不到可视的人工痕迹.


(a) News (b) Mobile

(c) Coastguard (d) Hall

图 8 原始视频在 H.264-E1 后的第 15 帧

为客观评价水印对视觉质量的影响，我们定义峰值信照比改变率($\Delta PSNR_r$)


(a) News (b) Mobile

(c) Coastguard (d) Hall

图 9 水印视频在 H.264-E1 后的第 15 帧

$$\Delta PSNR_r=\frac{PSNR'_{we-o}-PSNR_{oe-o}}{PSNR_{oe-o}}\times 100\% \quad (19)$$

其中 $PSNR'_{we-o}$表示原始含水印的视频在 H.264-E1 后与原始视频建立的 $PSNR$，而 $PSNR_{oe-o}$代表原始视频在 H.264-E1 后相对与原始视频建立的 $PSNR$. 通过对测试视频 $Y$ 分量的 $\Delta PSNR_r$ 计算，结果如图 10 所示. 水印嵌入后，5 个测试视频的 $\Delta PSNR_r$ 波动区间分别为[−16.40%，−11.17%]、[−17.34%，−10.88%]、[−12.07%，−8.21%]、[−11.17%，−3.30%]、[−14.86%，−10.69%]，均值分别为 −13.31%、−13.08%、−10.05%、−5.79%、−12.56%. 这些数据表明水印的嵌入导致视频在视觉质量上有一定程度的下降.

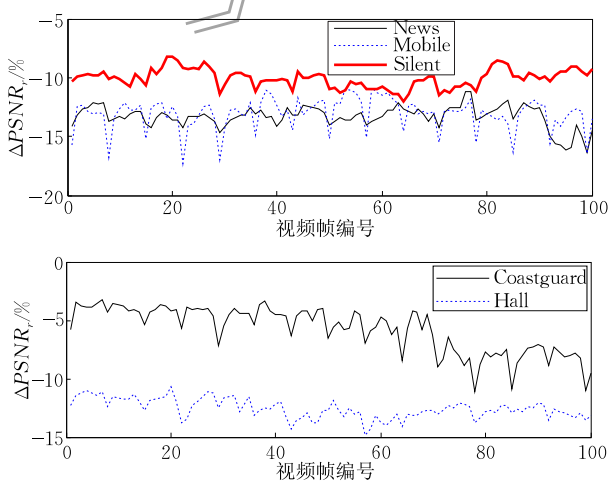



图 10 有/无水印的视频在 H.264-E1 后的 $PSNR$ 变化

然而，人眼的感知受诸多因素的影响，对图像误差的敏感度是相对的，导致基于 $PSNR$ 的度量结果难以与人眼观察的品质完全一致. 而文献[15]提出了一种基于 HVS 的结构相似度客观评价方法

(Structural Similarity Index,SSIM),度量结果与人眼观察的品质更一致,在广播电视中广为使用.*SSIM* 度量值在 0 到 1 的范围内,等于 0 时表明比较的对象完全不同,为 1 时说明比较对象完全一致.为进一步客观评价水印对视觉质量的影响,我们定义结构相似度改变量($\Delta SSIM$),如式(20)所示,

$$\Delta SSIM = SSIM'_{we-o} - SSIM_{oe-o} \tag{20}$$

其中符号 $SSIM'_{we-o}$ 与 $SSIM_{oe-o}$ 所代表的意思与式(19)类似.图 11 给出水印前后视频各帧在 *SSIM* 值上的变化,$\Delta SSIM$ 的波动范围从 −0.024 到 −0.002,根据 *SSIM* 的度量,由水印引入的视频视觉质量下降可以忽略.

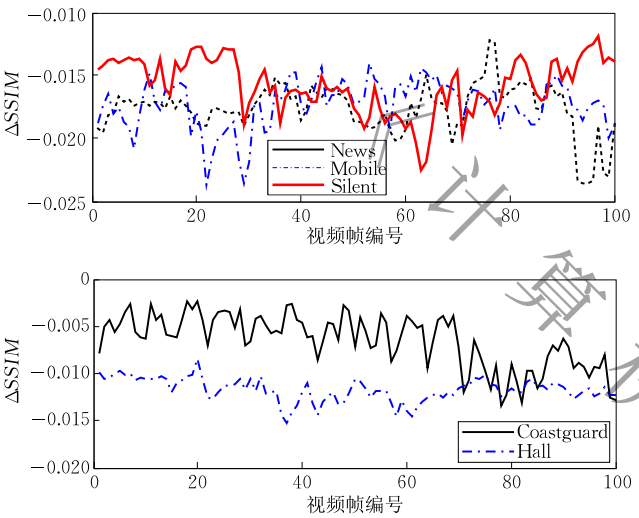



图 11　水印前后视频各帧在 *SSIM* 值上的变化

## 6.2　水印对码率的影响

本文通过定义比特率改变率($\Delta R_b$)来度量水印对码率的影响,如式(21)所示,

$$\Delta R_b = \frac{R'_b - R_b}{R_b} \times 100\% \tag{21}$$

其中$R'_b$、$R_b$ 分别表示始水印视频与原始视频经 H.264-E1 后的码率(各编码 200 帧).图 12 展示了

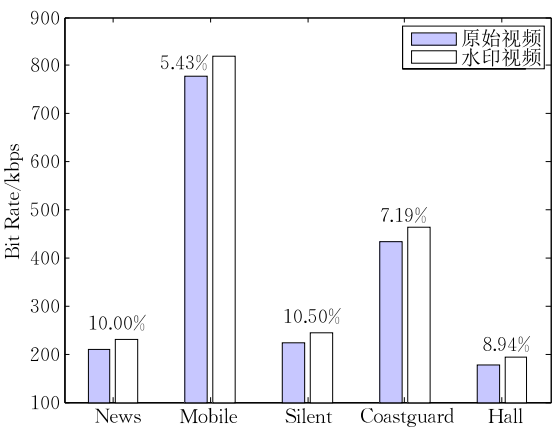



图 12　原始视频与原始水印视频经 H.264-E1 后的码率比较

水印嵌入视频前后码率的比较结果,水印导致码率平均增加 8.41%,原因来自两方面:(1)在水印嵌入算法中,对第一奇异值贡献率的量化主要是采用向前量化的方式,导致大部分数据块第一主元信息比例增加;(2)水印的嵌入破坏了原始数据块信息的自然流畅分布,信息冗余度较之前有所下降.

## 6.3　篡改与检测

在很多应用中,用户更关心的是视频内容或语义是否被改变,而非构成视频数据的数字微小改变.被保护的视频在 H.264-E1 后发布,能保持内容不变的应用途径可归纳为以下 3 种:(1)直接使用,即解码播放,不做任何数据处理;(2)为降低码率传播而解码、调整参数再编码;(3)根据特定应用需求而解码、图像增强、调整参数再编码.

为验证本文认证算法对保持内容处理的检测性能,我们执行了一系列实验.先设定 4.3 节涉及的两个检测阈值 $\tau_1 = 0.04$、$\tau_2 = 0.1$;检测效果如图 13 所示,表 1 给出了相关说明及检测结果.在图 13 中,(a1)～(a4)、(b1)～(b4)、(c1)～(c4)分别对应上述 3 种应用的检测效果.表 1 中的处理过程"H.264-E1＋13-缩放小(0.8×0.8)＋H.264-E2"表示原始视频加载水印后依次执行 H.264-E1 编/解码、编号为 13 的单项信号处理(第 2 节)、H.264-E2 变参再编码;此外,表 1 中"帧-B19"与"帧-I8"分别表示所对应的效果图为测试视频的第 19 帧与第 8 帧,且最后一次的编码的帧类型分别为 B 帧与 I 帧.通过将 *BER* 实验值与 $\tau_1$、$\tau_2$ 比较,可自动把当前测试帧判定为真实、基本可信、不可靠之一.

此外,我们还对视频空域内容的恶意篡改开展了实验,检测效果与相关说明分别如图 14、表 2 所示,其中恶意篡改是在上述保持内容处理的基础上(第(3)种应用途径)进行的,包括局部篡改与整帧替换.由于在篡改区域水印信息丢失,而且单个认证单元嵌入的是 1 bit 水印,所以相对于篡改区域局部的漏警率理论值是 0.5.再加上为消除由保持内容处理引入的噪声而采用的过滤措施,同时也消除了篡改区域被识别出的稀疏块,从而导致相对于篡改区局部的漏警率的实验值略高于 0.5(如图 14(a3)、(b3)).当篡改比例($p_T$)太小时,帧内容可能被判定为基本可信,此时要根据检测蒙版图进入人工干预过程.

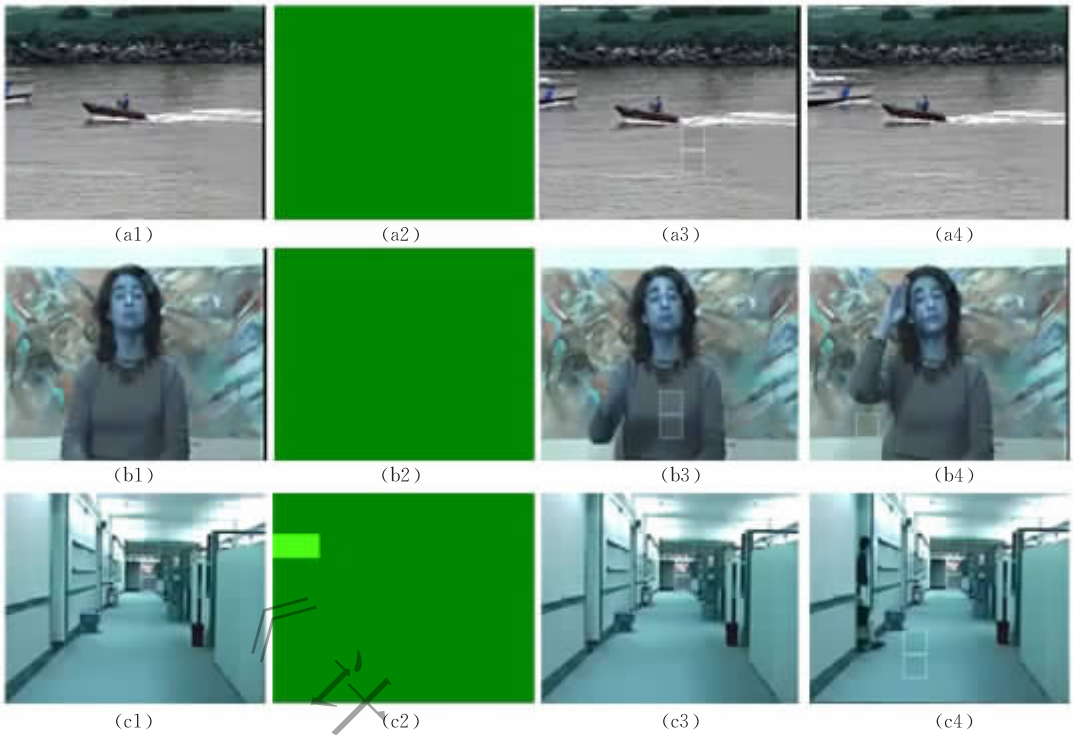



图 13 视频保持内容的处理与检测效果图(详细说明见表 1)

**表 1 视频保持内容处理的检测说明与结果($BER_{理论}=P_{fr|G\,理论}=0.04$; $\tau_1=0.04$, $\tau_2=0.1$)**

| 图号说明 | 处理过程 | 检测结果 | 图号说明 | 处理过程 | 检测结果 |
|---|---|---|---|---|---|
| (a1) Coast,帧-I8 | H.264-E1 | 对比参考帧 | (b3) Silent,帧-P13,检测蒙版图 | H.264-E1+H.264-E2 | $BER_{实验}=0.02$,图像真实 |
| (a2) Coast,帧-I8,检测二值图 | H.264-E1 | $BER_{实验}=0$,图像真实 | (b4) Silent,帧-B19,检测蒙版图 | H.264-E1+H.264-E2 | $BER_{实验}=0.01$,图像真实 |
| (a3) Coast,帧-P13,检测蒙版图 | H.264-E1 | $BER_{实验}=0.02$,图像真实 | (c1) Hall,帧-I9 | H.264-E1+13-缩放小(0.8×0.8)+H.264-E2 | 对比参考帧 |
| (a4) Coast,帧-P19,检测蒙版图 | H.264-E1 | $BER_{实验}=0$,图像真实 | (c2) Hall,帧-I9,检测二值图 | H.264-E1+13-缩放小(0.8×0.8)+H.264-E2 | $BER_{实验}=0.02$,图像真实 |
| (b1) Silent,帧-I8 | H.264-E1+H.264-E2 | 对比参考帧 | (c3) Hall,帧-B10,检测蒙版图 | H.264-E1+13-缩放小(0.8×0.8)+H.264-E2 | $BER_{实验}=0.0$,图像真实 |
| (b2) Silent,帧-I8,检测二值图 | H.264-E1+H.264-E2 | $BER_{实验}=0$,图像真实 | (c4) Hall,帧-P21,检测蒙版图 | H.264-E1+13-缩放小(0.8×0.8)+H.264-E2 | $BER_{实验}=0.02$,图像真实 |

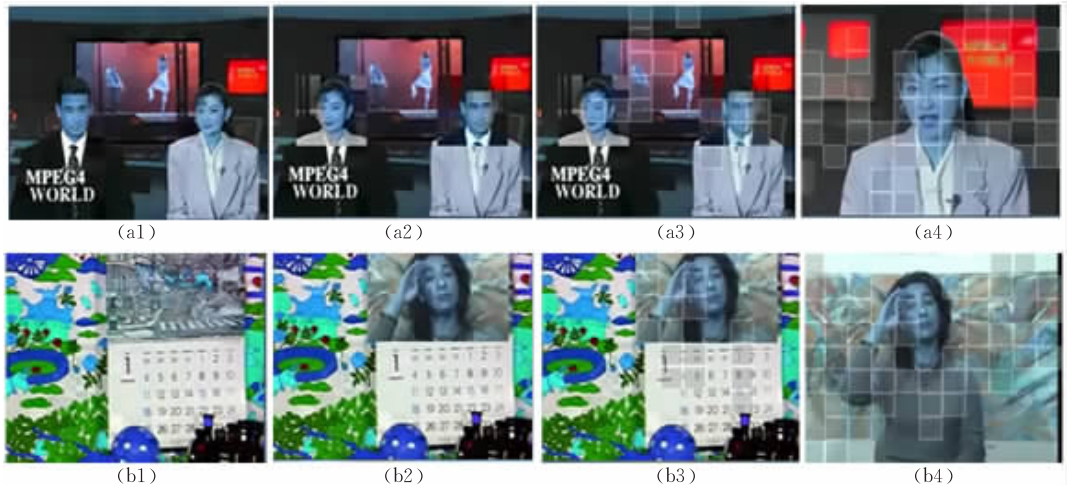



图 14 恶意篡改与检测效果图(详细说明见表 2)

**表 2　视频恶意篡改检测说明与结果($\tau_1=0.04,\tau_2=0.1$)**

| 图号说明 | 处理过程 | 检测结果 | 图号说明 | 处理过程 | 检测结果 |
|---|---|---|---|---|---|
| (a1) News,帧-B6 | H.264-E1 | 对比参考帧 | (b1) Mobile,帧-P13 | H.264-E1 | 对比参考帧 |
| (a2) News,帧-B6 | H.264-E1+局部篡改+6-滤波维纳(5×5)+H.264-E2 | 局部篡改图，$p_T=0.182$（篡改比例） | (b2) Mobile,帧-P13 | H.264-E1+局部篡改+11-相对增亮([0.4,1])+H.264-E2 | 局部篡改图，$p_T=0.242$ |
| (a3) News,帧-B6,检测蒙版图 | H.264-E1+局部篡改+6-滤波维纳(5×5)+H.264-E2 | $P_{r\|T\text{理论}}=0.875$，$P_{fa\|T\text{理论}}=0.086$，$P_{fr\|T\text{理论}}=0.066$；$P_{r\|T\text{实验}}=0.848$，$P_{fa\|T\text{实验}}=0.101$，$P_{fr\|T\text{实验}}=0.070$，$BER_{\text{实验}}=0.152$；图像不可信 | (b3) Mobile,帧-P13,检测蒙版图 | H.264-E1+局部篡改+11-相对增亮([0.4,1])+H.264-E2 | $P_{r\|T\text{理论}}=0.836$，$P_{fa\|T\text{理论}}=0.116$，$P_{fr\|T\text{理论}}=0.075$；$P_{r\|T\text{实验}}=0.848$，$P_{fa\|T\text{实验}}=0.172$，$P_{fr\|T\text{实验}}=0.080$，$BER_{\text{实验}}=0.152$；图像不可信 |
| (a4) News,帧-B16,帧替换,检测蒙版图 | H.264-E1+整帧替换+6-滤波维纳(5×5)+H.264-E2 | $p_T=1.0$，$P_{r\|T\text{理论}}=0.5$，$P_{fa\|T\text{理论}}=0.5$，$P_{fr\|T\text{理论}}=0.0$；$P_{r\|T\text{实验}}=0.535$，$P_{fa\|T\text{实验}}=0.465$，$P_{fr\|T\text{实验}}=0.0$，$BER_{\text{实验}}=0.465$ 图像不可信 | (b4) Mobile,帧-I25,帧替换,检测蒙版图 | H.264-E1+整帧替换+11-相对增亮([0.4,1])+H.264-E2 | $p_T=1.0$，$P_{r\|T\text{理论}}=0.5$，$P_{fa\|T\text{理论}}=0.5$，$P_{fr\|T\text{理论}}=0.0$；$P_{r\|T\text{实验}}=0.545$，$P_{fa\|T\text{实验}}=0.545$，$P_{fr\|T\text{实验}}=0.0$，$BER_{\text{实验}}=0.455$ 图像不可信 |

通过以上两类实验，验证了本文认证算法能有效过滤视频在实际应用中保持内容处理的情形，同时也说明对一定比例的恶意篡改有较好的区分度.受视频编码、信号处理、视频样本(数量、分辨率和纹理)等复杂因素的影响，表 1、表 2 中关于 $P_{r|T}$、$P_{fa|T}$、$P_{fr|T}$、$BER$ 各自的理论值与实验值存在偏差，但因偏差很小，正说明在第 2 节与第 3 节测试数据基础之上建立起的有关性能推导模型(5.2 节、5.3 节)的实用性.

### 6.4　性能比较

基于 DCT 系数的符号统计特性，文献[9-10]分别提出了适用于 H.264 视频内容认证的半脆弱水印算法；为进一步表明本文算法的实用性与优势，本节根据水印对视觉质量与码率的影响、水印的鲁棒性分别与二者进行比较.测试对象选取文献[9-10]中共同采用的 4 个 QCIF 格式标准视频序列(News、Tempete、Table、Mobile)，并设置编码参数为 $GOP$=IPPPPPPPPPI，$QP=28$.

水印对视觉质量与码率影响的比较结果分别如表 3、表 4 所示.表 3 中的数据说明文献[9-10]的算法对视频的视觉质量几乎没影响，而本文算法在人眼难以察觉的范围内使信号质量略微失真，$PSNR$ 平均下降 2.82dB.根据表 4 数据，水印导致码率的增加平均值分别为：8.15%(本文)、7.14%(文献[9])和 3.84%(文献[10])，相比之下，本算法的码率性能略低，但不影响其实用性.

**表 3　基于 *PSNR* 改变量的性能比较**　(单位：dB)

| | News | Tempete | Table | Mobile | (Avg.) |
|---|---|---|---|---|---|
| 文献[10] | −0.02 | −0.02 | −0.01 | −0.01 | −0.015 |
| 文献[9] | −0.05 | −0.03 | −0.03 | −0.02 | −0.033 |
| 本文 | −4.50 | −2.49 | −0.77 | −3.52 | −2.820 |

**表 4　基于比特率改变率($\Delta R_b$)的性能比较**　(单位：%)

| | News | Tempete | Table | Mobile | (Avg.) |
|---|---|---|---|---|---|
| 文献[10] | 7.35 | 1.12 | 4.58 | 2.31 | 3.84 |
| 文献[9] | 14.27 | 3.66 | 6.79 | 3.82 | 7.14 |
| 本文 | 6.99 | 7.58 | 6.77 | 11.28 | 8.15 |

由于半脆弱水印算法的主要目的是能区分保持内容的处理与恶意攻击，所以算法的鲁棒性成为了更重要的性能指标，接下来我们采用基于初始篡改矩阵的比特错误率($BER_T$)与文献[9-10]进行比较.

$$BER_T=\frac{256\times\sum_{i=1}^{m_1/16}\sum_{j=1}^{m_2/16}t_{i,j}^0}{m_1\times m_2} \tag{22}$$

鲁棒性比较结果如表 5 所示.为了消除因视频亮度的绝对增/减而导致对水印稳定性的影响，本文在水印嵌入时采用了标准化处理(4.2 节，算法 1)，然而标准化处理又引起极少数数据块的特征在检测时发生改变现象，所以在表 5 中出现了在无攻击的情况下本文算法的 $BER_T$ 不等于 0 情形.表 5 中对常规信号处理采用了均值为 0，标准差分别为 1 和 2 的高斯白噪声，以及亮度绝对增加 10 和 20 的操作.由于文献[9-10]为编码域算法，为了公平比较，“一

次压缩”处理对应本文原始域算法的一次编码($GOP$=IPPPPPPPPPI,$QP$=26),而其对应文献[9-10]的如下处理过程:在编码过程嵌入水印($GOP$=IPPPPPPPPPI,$QP$=28),并按 $QP$=26 重编码;“二次压缩”为一次压缩后的解码与再编码($GOP$=IBBBPBBBI,$QP$=28).

**表 5 基于 $BER_T$ 的鲁棒性比较** (单位:%)

| 视频 | News | | | Tempete | | | Table | | | Mobile | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 本文 | [10] | [9] | 本文 | [10] | [9] | 本文 | [10] | [9] | 本文 | [10] | [9] |
| 无攻击 | 1.42 | 0 | 1.41 | 0.84 | 0 | 2.83 | 1.11 | 0 | 4.42 | 1.39 | 0 | 2.02 |
| 噪声(0,$\sigma$=1) | 1.42 | 5.81 | 5.72 | 0.84 | 7.42 | 7.54 | 1.11 | 8.56 | 9.70 | 1.39 | 7.19 | 7.41 |
| 噪声(0,$\sigma$=2) | 1.42 | 5.98 | 6.46 | 0.84 | 8.76 | 10.51 | 1.11 | 10.57 | 11.18 | 1.39 | 9.23 | 10.71 |
| 绝对增亮(+10) | 1.42 | 4.66 | 5.39 | 0.84 | 6.03 | 6.94 | 1.11 | 9.21 | 9.63 | 1.39 | 6.60 | 6.46 |
| 绝对增亮(+20) | 1.42 | 6.28 | 6.33 | 0.84 | 6.79 | 7.61 | 1.11 | 9.59 | 9.70 | 1.39 | 6.78 | 6.94 |
| 一次压缩 | 9.63 | 5.46 | 8.82 | 5.02 | 17.72 | 14.41 | 18.04 | 13.73 | 12.05 | 7.47 | 16.36 | 14.81 |
| 二次压缩 | 17.87 | **94.21** | **50.63** | 10.07 | **93.84** | **51.26** | 23.21 | **93.85** | **51.39** | 12.28 | **93.47** | **50.20** |

根据表 5 中的数据,本文算法在常规信号处理下的鲁棒性全优于文献[9-10].由于噪声强度较小,且本文水印算法在嵌入和提取时采用了标准化处理,所以算法的 $BER_T$ 值在无攻击与有攻击(表中 4 种常规信号处理)下保持了一致.针对一次压缩处理,本文算法与文献[9-10]各有优劣,而对二次压缩,本文算法有着优势显著.由于二次压缩改变了 $GOP$ 参数,使的文献[9-10]的检测算法不再与水印嵌入时同步;因文献[9]为 1 bit 水印嵌入,检测未同步时 $BER_T$ 理论值为 0.5;而文献[10]在同一宏块数据中采用相同算法分别嵌入 4 个独立的水印,如此做法的鲁棒性不及 1 bit 嵌入,但能降低漏检率,在检测未同步时 $BER_T$ 理论值为 $1-(1/2)^4=0.9375$.

通过性能比较表明,本文算法在对视觉质量与码率影响符合实用的情况下,鲁棒性优于文献[9-10],尤其是在二次压缩条件下本文算法可以区分保持内容的处理与恶意篡改,而文献[9-10]的算法都不能实现.

## 7 结 论

视频以压缩形式发布后,往往会经历解码、常规信号处理和修改 $GOP$ 及 $QP$ 参数的二次压缩,再进入应用领域,而经历此过程的处理后,视频的内容(语义)未发生改变.然而现有的视频内容认证水印算法面对如此二次压缩的处理过程,检测是失败的.本文基于量化调制奇异值贡献率的方法提出了一种新颖的视频内容认证水印算法,并通过以下方式来提高算法的各项性能:(1)在视频 $Y$ 分量的分块 2 级 DWT 高频域,基于能量关系建立鲁棒的特征码;(2)通过量化调制第一奇异值贡献率的方法,实现水印抗二次压缩的鲁棒性;(3)在水印嵌入的同时,通过保持其余奇异值贡献率在调整前后相互间的比例关系不变,能提高水印的透明性;(4)根据密钥及特征水印的交叉嵌入方法来提高算法的安全性;(5)根据邻域特征能有效降低检测的虚警率.本文算法能有效抵制密码分析攻击、拼贴攻击、量化攻击及针对性的伪造攻击;实验表明,提出的算法水印透明性好,对码率影响小,并且在二次 H.264 变参压缩及期间的常规信号处理条件下,能有效区分保持内容的处理与恶意篡改攻击.

## 参 考 文 献


[1] Zhao Y, Wang S, Zhang X, et al. Robust hashing for image authentication using Zernike moments and local features. IEEE Transactions on Information Forensics and Security, 2013, 8(1): 55-63

[2] Wen Zhen-Kun, Gao Jin-Hua, Zhu Ying-Ying, et al. Video perceptual hashing fusing spatiotemporal change detection. Acta Electronica Sinica, 2014, 42(6): 1163-1167(in Chinese)
(文振焜, 高金花, 朱映映等. 融合时空域变化信息的视频感知哈希算法研究. 电子学报, 2014, 42(6): 1163-1167)

[3] Su P C, Chen C C, Chang H M. Towards effective content authentication for digital videos by employing feature extraction and quantization. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(5): 668-677

[4] Queluz M P. Authentication of digital images and video: Generic models and a new contribution. Signal Processing: Image Communication, 2001, 16(5): 461-475

[5] Doerr G, Dugelay J L. A guide tour of video watermarking. Signal Processing: Image Communication, 2003, 18(4): 263-282

[6] Chen H Y, Zhu Y S. A robust video watermarking algorithm based on singular value decomposition and slope-based embedding technique. Multimedia Tools and Applications, 2014, 71(3): 991-1012

[7] Wang Xiao-Jing, Yang Gao-Bo, Zhu Ning-Bo. Content authentication of H.264/AVC video stream based on semi-fragile watermarking. Journal on Communications, 2009, 30(11): 71-78(in Chinese)
(王小静, 杨高波, 朱宁波. 基于半脆弱水印的 H.264/AVC

视频流的内容级认证. 通信学报，2009，30(11)：71-78)

[8] Lin Zhi-Gao，Sun Tan-Feng，Jiang Xing-Hao. A content level watermarking scheme for H.264/AVC video authentication in VLC domain. Journal of Shanghai Jiaotong University，2011，45(10)：1531-1535(in Chinese)
(林志高，孙锬锋，蒋兴浩. 基于 VLC 域的 H.264/AVC 视频流内容级认证水印算法. 上海交通大学学报，2011，45(10)：1531-1535)

[9] Xu D W，Wang R D，Wang J. A novel watermarking scheme for H.264/AVC video authentication. Signal Processing：Image Communication，2011，26：267-279

[10] Farfoura M E，Horng S J，Guo J M，Al-Haj A. Low complexity semi-fragile watermarking scheme for H.264/AVC authentication. Multimedia Tools and Applications，2016，75(13)：7465-7493

[11] Wilkinson J H. The Algebraic Eigenvalue Problem. Oxford，UK：Clarendon Press，1965

[12] Dai Y W，Thiemert S，Steinebach M. Feature-based watermarking scheme for MPEG-I II video authentication// Proceedings of the SPIE Security，Steganography，and Watermarking of Multimedia Contents. California，USA，2004，5036：325-335

[13] Wu Chun-Guo，Liang Yan-Chun，Sun Yan-Feng，et al. On the equivalence of SVD and PCA. Chinese Journal of Computers，2004，27(2)：286-288(in Chinese)
(吴春国，梁艳春，孙延风等. 关于 SVD 与 PCA 等价性的研究. 计算机学报，2004，27(2)：286-288)

[14] Wu Jin-Hai，Lin Fu-Zong. Image authentication based on digit al watermarking. Chinese Journal of Computers，2004，27(9)：1153-1161(in Chinese)
(吴金海，林福宗. 基于数字水印的图像认证技术研究. 计算机学报，2004，27(9)：1153-1161)

[15] Wang Z，Bovik A C，Sheikh H R，Simoncell E P. Image quality assessment：From error visibility to structural similarity. IEEE Transactions on Image Processing，2004，13(4)：600-612

**FU Jian-Jing**，born in 1976，Ph. D.，associate professor. His research interests include digital watermarking，digital content security，etc.

**CHEN De-Ren**，born in 1951，professor，Ph. D. supervisor. His research interests include information security，electronic commerce，data mining，etc.



**Background**

With the rapid development of the multimedia technology and video editing software，video data can be easily altered and forged，which makes difficult to distinguish the true from the false. When it comes to the court evidence，medical identification，military intelligence，national security and other fields，the video information tampering may cause the grave consequences. How to effectively identify the authenticity of the digital video content has received extensive attention.

At present，the technology used for digital video content authentication is perceptual hash and digital watermarking technology. As video perceptual hash is used for content authentication，there is a need to the extra overhead to transfer the digest information，and it cannot locate the specific location of the content tampering，so its application in content authentication has a lot of limitations. However，for the existing algorithms based on semi-fragile video watermarking，its watermark can survive only after conventional signal processing or recompression condition，but cannot be detected，when the tested video is processed by conventional signal processing and recompression.

Besides，in practical application，in order to achieve the video theft，malicious tampering，no malicious signal processing，and reduce the bandwidth，the video is necessary to be performed secondary compression，which includes video decoding，signal processing，and encoding again. But during the secondary encoding compression，it is very likely to modify parameters of the *GOP*. It is worth noting that after the processing of above procedure，the content of the video (semantic) has not changed. However，the existing watermarking algorithm for video content authentication is failure to detect watermark when video is subjected to such secondary compression as above. Therefore，how to determine the authenticity of video content is an urgent problem to be solved under the condition of the secondary compression.

As for it，this paper proposes a novel watermarking algorithm for video content authentication. The experimental results show that the proposed algorithm can effectively distinguish between video content-preserving processing and malicious tampering and forgery under the condition of the secondary compression of H.264/AVC，and other performance indicators also meet the practical needs.

This work is supported by the National Natural Science Foundation of China (No. 61502415)，the Zhejiang Provincial Natural Science Foundation (Nos. LY18F020003，LY15F020016)，and the Public Welfare Technology Application Research Project of Zhejiang Province (No. 2015C31110).